Reviewer’s report

Title: Analysis of influenza transmission in the households of primary and junior high school students during the 2012/13 influenza season in Odate, Japan

Version: 1  Date: 16 October 2014

Reviewer: Jonathan D Sugimoto

Reviewer’s report:

Summary
The authors report the results of a survey of the households of primary school and junior high school students in Odate City, Japan. The survey assessed self-reported influenza-like illness among the members of households. The authors use the relative onset times for illness among household members to estimate the serial interval distribution, as well as the effects of demographic characteristics and interventions on the SI (e.g., seasonal vaccination and antiviral drugs). The authors estimates the mean (standard deviation) for the SI as 2.8 (3.6) days.

General Comments
The study that the authors are reporting is quite interesting and pertinent for informing the ongoing discussion of the role of households, especially those of school-age children, facilitating the community transmission of influenza. Though the authors’ study provides some interesting and potentially important results, some key points and details about the data collection and analytic methods still need to be clarified.

Specific Comments

Major Compulsory Revisions

1. Line 127: The exclusion of the household members who developed symptoms on the same day as the index/primary case is likely leading to a bias to toward more “transmission events” being attributed to occurring “from PS or JH students”, given that PS and JH students were defined a priori as the primary case for a household. Given that the authors appear to have information about the occurrence and timing of the onset of influenza illness all members of the household, please provide a very strong rationale for why you chose to exclude these individuals from the analysis. I understand the reasoning behind the choice to exclude the individuals whose onset was greater than 7 days after the onset day for the primary/index case, so there is no need to justify that choice.

2. Please be very explicit about how you determined who-infected-whom. This is a key assumption of your model for how transmission occurred in these households, so I would like to see a very clear description of how this choice was made.
3. As mentioned below under a “Minor Essential Revisions” comments, the fact that this household survey did not collect information about the size of the household (alternatively, or a list of the membership) should be stated directly in the methods, along with a reason for why this information was not obtained. This information is central to understanding transmission in the household.

4. For the paragraphs describing the estimation of the SI distribution and covariate effects on this parameter, please integrate the specification of the statistical methods used and the sources for these methods (e.g., the Hong Kong study) into the first paragraph.

5. Other than a reference to why the SAR was not calculated, there is no discussion of the potential effects (biases etc…) of the very low response/coverage rate (12.4% of eligible households). Please provide some discussion of this limitation in the appropriate section.

6. Though the authors provide a valid reason for deciding not to calculate the SAR, i.e., selection bias in the self-reporting of illness, reporting this information, if estimable, is still a useful exercise. SAR estimates based upon this type of study design are routinely reported in the literature, and methods exist to at least partially account for this selection bias issue (see the references listed at the end of the review). The authors might consider reporting this information, or changing the reasoning behind not reporting to include a statement to the effect of “data on the size of the household was not collected, thereby precluding the estimation of the SAR.”

7. The relationship between the SI and the length of time spent at home is likely to be partially confounded by differences in the amount time that other household members were exposed to children who stayed home a relative long time versus those that only stayed home for a relatively shorter time. Even if the length of the incubation period was the same for all of the PS and JH students in the study, a longer time spent at home would lead to an exposure of other household members to this study over a longer period, thereby prolonging the mean duration of the SI for these households, relative to those where students were not at home as long. Please make this potential for confounding clear in your interpretation of this particular result.

Minor Essential Revisions

1. The second sentence of the abstract needs to be reworded for clarity.

2. In different locations throughout the manuscript, the year during which the study was conducted is presented alternatively as 2011-2012 or 2012-2013. Please ensure that this specification is consistent throughout the manuscript.

3. Line 115-116 (Methods Section): Please clarify whether or not the surveillance system records were linked at the individual-level to the results of the household survey. Specify how they may have been related. For example, were all of the cases detected by the household survey automatically reported to the surveillance system?

4. Data analysis subsection of the Methods Section: In general, the terms
“primary” and “index” case have two similar but different meanings. A primary case refers to the actual first case to occur in a household during a defined period of time. An index case is the case that leads to a household being brought to the attention of a surveillance or monitoring system. From the description of your study, the term primary case would seem more appropriate.

5. Just as a note, the use of the term “secondary case” implies that these cases were infected by a primary case in the household. The analytic methods described make this assumption, though this assumption was almost certainly violated for some of these households.

6. The second clause of the first sentence of the data analysis subsection of the Methods Section is unclear. Please clarify.

7. Line 180 of the Results section the number of households reporting at least one influenza case is not consistent with the 43% figure reported in line 171.

8. Please specify the case definition for influenza used for the household survey. Such a definition is provided for the surveillance system, but there doesn’t seem to be a description of how influenza illness was defined for the household survey.

9. Line 134-135: Please clarify why you a priori defined primary cases to be PS and JH students.

10. Line 207-208: Please clarify the meaning of the sentence beginning with “Referenced to fathers, …”

11. Line 249-250. This conclusion doesn’t appear to be completely warranted given the available data and analytic methods, as described. Without any information about the number of people who were at-risk for influenza illness in the household (for example, the entire membership of the household), it is hard to see how the authors can conclude anything about the nature of the “probability of household transmission”. To make this type of statement, the authors would need to have estimated at minimum a parameter related to the probability of transmission, such as the secondary attack rate for the household. Please amend this statement to better reflect the appropriate interpretation of this study’s results.

12. Line 262-263: Please clarify the meaning of this sentence.

13. Line 287-289: The fact that the size of the household was not collected for each of the 352 surveyed households (if indeed that is the appropriate interpretation of this sentence) should be included in the materials and methods section, because this strongly affects how at least I interpret this study’s results. If my interpretation of this statement is incorrect and instead the authors did collect household size information, then the authors might consider estimating and reporting the SAR, possibly even estimating the probability of transmission in a more rigorous manner (for example, see Yang et al. 2006 or Cauchemez et al 200X).

14. Please present 95% confidence intervals in place of means and standard deviations/errors.

Discretionary Revisions
None

Bibliographic References.


Both of these references are from two of the leading biostatisticians developing methods for the analysis of household data. These are not the primary references for the models that they use, but you might find them useful as starting point.

Level of interest: An article of importance in its field

Quality of written English: Acceptable

Statistical review: No, the manuscript does not need to be seen by a statistician.

Declaration of competing interests:

None